

Simulation of multiple partonic interactions in **Herwig++**

Manuel Bähr, Stefan Gieseke

*Institut für Theoretische Physik
Universität Karlsruhe, 76128 Karlsruhe, Germany*

Michael H. Seymour

*School of Physics and Astronomy, University of Manchester; and
Physics Department, CERN, CH-1211 Geneva 23, Switzerland*

ABSTRACT: In this paper we describe a new model of multiple partonic interactions that has been implemented in **Herwig++**. Tuning its two free parameters we find a good description of CDF underlying event data. We show extrapolations to the LHC.

KEYWORDS: Hadronic Colliders, QCD, Jets, Phenomenological Models, Underlying Event.

Contents

1. Introduction	1
2. Details	4
2.1 Eikonal model	4
2.1.1 Different scattering types	6
2.2 Monte Carlo implementation	6
2.2.1 Parton extraction	7
3. Results	9
3.1 Tuning and Tevatron results	9
3.2 PDF uncertainties	11
3.3 LHC expectation	13
4. Conclusions	15
A. Forced splitting: implementation in the event record	16
B. Model parameters	17
C. Systematic errors in the low p_T region	17

1. Introduction

With the advent of the Large Hadron Collider (LHC) in the near future it will become increasingly important to gain a detailed understanding of all sources of hadronic activity in a high energy scattering event. An important source of additional soft jets will be the presence of the underlying event. From the experimental point of view, the underlying event contains all activity in a hadronic collision that is not related to the signal particles from the hard process, e.g. leptons or missing transverse energy. The additional particles may result from the initial state radiation of additional gluons or from additional hard (or soft) scatters that occur during the same hadron–hadron collision. Jet measurements are particularly sensitive to the underlying event because, although a jet’s energy is dominated by the primary hard parton that

initiated it, jet algorithms inevitably gather together all other energy deposits in its vicinity, giving an important correction to its energy and internal structure.

In standard Monte Carlo event generators, like **Herwig**(++) [1–6], **PYTHIA** [7, 8] or **SHERPA** [9], additional gluons from initial and final state radiation are generated with the help of parton shower algorithms, possibly supplemented by multijet matrix elements [10, 11]. Therefore, we tend to attribute these to the hard process rather than to the underlying event. On top of that, the underlying event is simulated as some additional hadronic activity. The simplest way to do so is the so-called UA5 model [12], which has been the default underlying event model in **Herwig**++ for a long time. Here, additional (soft) hadronic activity is generated as a number of additional clusters are generated flat in rapidity with an exponential transverse momentum distribution. See [5] for more details. These clusters eventually give the required additional activity of soft hadrons.

Another variant, which has been far more successful in the description of recent collider data, was formulated as a sequence of more-or-less independent parton interactions. In contrast to the UA5 model this model is capable of describing the jet-like structure of the underlying event. In its initial formulation [13] there were no parton showers invoked. Later variants of this model also contain full parton showers [14, 15]. The additional scatters in these models are always modelled as simple QCD $2 \rightarrow 2$ scattering as long as the scattering contains a hard jet of at least a few GeV. Soft, more forward scattering may also be modelled but requires a unified description of perturbative and non-perturbative scattering, as in the dual parton model [16–18], which had been implemented into the event generator **PHOJET** [19]. Another model is the simple extrapolation of the transverse momentum distribution of hard jets in QCD processes down to zero p_T [20]. Such a modelling of soft interactions will also allow us to describe minimum bias events. These are dominated by soft, forward scatterings and diffractive production of particles during the hadron–hadron scattering event.

Experimentally, there has been strong evidence for the presence of multiple partonic interactions already at the CERN ISR through the measurement of a momentum imbalance in multijet events [21]. The idea for this measurement is that multiple pairs of jets, two in this case, will appear to be balanced in transverse momentum if they have been created in different back-to-back events rather than a single multijet event. Similar observations of double parton scattering [22] have been made at the Tevatron [23, 24]. Nowadays, the clearest observation has been made in $\gamma+3$ jet events at CDF [25]. In addition to this clear evidence for the presence of multiple interactions in hadronic collisions, the only sensible description of the final state of such events can be made with detailed Monte Carlo modelling, based on this ansatz. The most detailed measurements of the properties of the underlying event as well as their implications for Monte Carlo models are described in [26, 27].

Understanding minimum bias interactions and the underlying event are very

important for many aspects of LHC physics. Particularly in high luminosity runs, every triggered hard event in one bunch crossing will be accompanied by additional interactions among other protons from the same bunch. These are predominantly minimum bias interactions and will give some additional activity in the detectors. There are already detailed plans for the measurement of the underlying event in ATLAS [28] and CMS [29, 30]. The presence of the underlying event is important whenever measurements at the LHC will be based on the measurements of the properties of jets, like e.g. their energy. The determination of the so-called jet energy scale is known to be improved when a reasonable modelling of the underlying event is included in the analysis. A good example for this is the measurement of the top mass [31]. Implications for the central jet veto in vector boson fusion processes have been addressed in detail in [32].

In this paper we want to focus on the description of the hard component of the underlying event, which stems from additional hard scatters within the same proton. Not only does this model give us a simple unitarization of the hard cross section, it also allows to give a good description of the additional substructure of the underlying events. It turns out that most activity in the underlying event can be understood much better in terms of hard minijets. We therefore adopt this model, based on the model **JIMMY** [14, 15], also for our new event generator **Herwig++** [5]. We will describe the basic implementation of the model and its parameters and study some important implications for jet final states. Thus far, we do not consider a description beyond multiple hard interactions. An extension of our model towards softer interactions along the lines suggested in [20] is planned and will also allow us to describe minimum bias interactions.

Improvements of the underlying event description have also been implemented in other event generators. A completely new formulation of the interleaving of underlying hard scatterings with the parton shower has been introduced with the latest versions of **PYTHIA** [7, 8, 33, 34]. A model very similar to the multiple interaction model in **PYTHIA** has been implemented in **SHERPA** [35]. A new approach, based on k_T -factorization [36–38] has been introduced and studied in [39]. An important issue, which has been addressed in [33] is the relation between the charged particle multiplicity and the average transverse momentum in the underlying event. The relation between these observables in the transverse region of jet events may point us towards the right colour correlations of the different hard scatters [40]. We want to point out that the organization of colour lines adopted in our model differs significantly from that in **PYTHIA**. In this paper we would like to focus on the details of the implementation, validation and tuning on Tevatron data and some predictions for the LHC.

This paper is organized as follows: In Sect. 2 we briefly review the theoretical motivation for multiple interactions and describe all details that are relevant for our Monte Carlo implementation. In Sect. 3 we discuss the parameters of our model and

perform a fit to current Tevatron data. Taking this as a starting point, we make predictions for the most important final state observables at the LHC. Furthermore, we discuss implications of the intrinsic uncertainties of parton distribution functions for the underlying event observables. In Sect. 4 we draw some conclusions and give an outlook to future work. Some more model details will be described in Appendices.

2. Details

The starting point for thinking about multiple interactions is the observation that the cross section for QCD jet production may exceed the total pp or $p\bar{p}$ cross section already at an intermediate energy range and eventually violates unitarity. For example, for QCD jet production with a minimum p_T of 2 GeV this already happens at $\sqrt{s} \sim 1$ TeV. This p_T cutoff should however be large enough to ensure that we can calculate the cross section using pQCD. The reason for the rapid increase of the cross section turns out to be the strong rise of the proton structure function at small x , since the x values probed decrease with increasing centre of mass energy. This proliferation of low x partons may lead to a non-negligible probability of having more than one partonic scattering in the same hadronic collision. This is not in contradiction with the definition of the standard parton distribution function as the *inclusive* distribution of a parton in a hadron, with all other partonic interactions summed and integrated out. It does, however, signal the onset of a regime in which the simple interpretation of the pQCD calculation as describing the only partonic scattering must be unitarized by additional scatters.

In principle, predicting the rate of multi-parton scattering processes requires multi-parton distribution functions, about which we have almost no experimental information. However, the fact that the standard parton distribution functions describe the inclusive distribution gives a powerful constraint, which we can use to construct a simple model.

2.1 Eikonal model

The eikonal model introduced in Refs. [14, 41, 42] derives from the assumption that at fixed impact parameter b , partons undergo independent scatters with mean number

$$\langle n(b = |\mathbf{b}|, s) \rangle = \int d^2\mathbf{b}' \int_{p_T^{\min 2}} dp_T^2 \sum_{ij} \frac{1}{1 + \delta_{ij}} \frac{d\hat{\sigma}_{ij}(x_1\sqrt{s}, x_2\sqrt{s}, p_T^2)}{dp_T^2} \otimes G_{i/h_1}(x_1, \mathbf{b} - \mathbf{b}', \mu^2) \otimes G_{j/h_2}(x_2, \mathbf{b}', \mu^2), \quad (2.1)$$

where $d\hat{\sigma}$ is the differential partonic cross section for QCD 2→2 scattering and $G(x, \mathbf{b}, \mu)$ are parton densities representing the average number of partons with a given momentum fraction x and transverse coordinate \mathbf{b} . \otimes denotes the convolution integrals in longitudinal momentum fractions x_1, x_2 . By further assuming a

factorization of the x and \mathbf{b} dependence in G , namely

$$G(x, \mathbf{b}, \mu^2) = f(x, \mu^2) \cdot S(\mathbf{b}), \quad (2.2)$$

where $f(x, \mu^2)$ is the conventional parton distribution, Eq. (2.1) can be written as

$$\begin{aligned} \langle n(b, s) \rangle &= \sigma^{\text{inc}}(s; p_T^{\min}) \cdot \int d^2\mathbf{b}' S_{h_1}(\mathbf{b}') S_{h_2}(\mathbf{b} - \mathbf{b}') \\ &= A(b) \cdot \sigma^{\text{inc}}(s; p_T^{\min}). \end{aligned} \quad (2.3)$$

In Eq. (2.3) σ^{inc} denotes the inclusive cross section to produce a pair of jets (partons) with $p_T > p_T^{\min}$ and is given by the standard perturbative calculation, whereas $A(b)$ describes the overlap of the partons in the colliding hadrons. We model the impact parameter dependence of partons in a hadron, $S(\mathbf{b})$, by the electromagnetic form factor,

$$S_{\bar{p}}(\mathbf{b}) = S_p(\mathbf{b}) = \int \frac{d^2\mathbf{k}}{2\pi} \frac{e^{i\mathbf{k}\cdot\mathbf{b}}}{(1 + \mathbf{k}^2/\mu^2)^2}, \quad (2.4)$$

where μ is the inverse hadron radius. This leads to

$$A(b) = \frac{\mu^2}{96\pi} (\mu b)^3 K_3(\mu b), \quad (2.5)$$

where $K_3(x)$ is the modified Bessel function of the third kind. We do not fix μ at the value determined from elastic ep scattering, but rather treat it as a free parameter, because the spatial parton distribution is assumed to be similar to the distribution of charge, but not necessarily identical.

The assumption that different scatters are uncorrelated leads to the Poissonian distribution for the number of scatters, k , at fixed impact parameter,

$$\mathcal{P}_k(\langle n \rangle) = \frac{\langle n \rangle^k}{k!} e^{-\langle n \rangle}. \quad (2.6)$$

The cross section for having exactly n scatters with individual cross section σ^{inc} , using this assumption is

$$\sigma_n(\sigma^{\text{inc}}) = \int d^2b \mathcal{P}_n(A(b) \cdot \sigma^{\text{inc}}) = \int d^2b \frac{(A(b) \cdot \sigma^{\text{inc}})^n}{n!} e^{-A(b) \cdot \sigma^{\text{inc}}}. \quad (2.7)$$

The probability of having n scatters in an event, given that there is at least one, is then

$$P_{n \geq 1}(\sigma^{\text{inc}}) = \frac{\int d^2b \mathcal{P}_n(A(b) \cdot \sigma^{\text{inc}})}{\int d^2b \sum_{k=1}^{\infty} \mathcal{P}_k(A(b) \cdot \sigma^{\text{inc}})} = \frac{\sigma_n(\sigma^{\text{inc}})}{\sigma_{\text{hard}}(\sigma^{\text{inc}})}. \quad (2.8)$$

Equation (2.8) is used as the basis of the multi-parton scattering generator for events in which the hard process is identical to the one used in the underlying event, i.e. QCD $2 \rightarrow 2$ scattering. For distinct scattering types a modification is used, as described in the next section.

2.1.1 Different scattering types

Following the assumption of independent additional scatterings the cross section for two distinct scattering types a and b with the respective multiplicities k and m can be written as

$$\begin{aligned}\sigma_{k,m}(\sigma_a, \sigma_b) &= \int d^2b \mathcal{P}_k(A(b)\sigma_a) \mathcal{P}_m(A(b)\sigma_b) \\ &= \int d^2b \frac{(A(b)\sigma_a)^k}{k!} e^{-A(b)\sigma_a} \frac{(A(b)\sigma_b)^m}{m!} e^{-A(b)\sigma_b}.\end{aligned}\quad (2.9)$$

For small signal cross sections σ_b , the exponential can be approximated by unity. Using Eq. (2.9) the probability of having k events of type a in the presence of exactly one of type b is

$$\begin{aligned}P_k &= \frac{\sigma_{k,1}}{\sum_{\ell=0}^{\infty} \sigma_{\ell,1}} \approx \frac{\int d^2b \mathcal{P}_k(A(b)\sigma_a) \cdot A(b)\sigma_b}{\int d^2b A(b)\sigma_b} \\ &= \int d^2b \mathcal{P}_k(A(b)\sigma_a) \cdot A(b).\end{aligned}$$

This can then be rewritten to avoid the extra factor $A(b)$ in the form,

$$P_{n=k+1} \approx \frac{n}{\sigma_a} \int d^2b \mathcal{P}_n(A(b)\sigma_a). \quad (2.10)$$

Here n is the total number of scatters, i.e. there is one of type b and $n-1$ of type a . It is worth noting that the fact that we have ‘triggered on’ a process with a small cross section leads to a bias in the b distribution and hence a higher multiplicity of additional scatters than in the pure QCD $2 \rightarrow 2$ scattering case.

Equation (2.10) can be used to describe underlying event activity under rare signal processes as well as jet production in the underlying event simulated under high p_T jet production as signal process. In the latter case the assumption of distinct scattering processes may not be fulfilled. One can show that in that case the m th scatter of type a that is also of type b should be rejected with probability $1/(m+1)$.

2.2 Monte Carlo implementation

The model introduced so far is entirely formulated at the parton level. However, an event generator aims for a full description of the event at the level of hadrons. This implies that the implementation of multi-parton scattering must be properly connected to the parton shower and hadronization models, a few details of which we discuss in the following. We give more technical details of the way in which the multiple scattering is represented in the event record, and of how to access the model parameters, in Appendices A and B respectively.

Event generation starts with the sampling of the hard process according to its matrix element and the parton densities. After that the parton shower evolves the

final state partons from the scale of the hard interaction down to a cutoff scale that is of the order of the confinement scale, but large enough to ensure that we remain within the perturbative regime. The incoming partons are evolved backwards to higher values of x and decreasing μ^2 . The initial- and final-state parton showers in **Herwig++** are performed using the coherent branching algorithm of Ref. [43], which is based on the original coherent shower algorithm of Refs. [44–46]. After the initial-state shower has terminated, the incoming partons are extracted out of the beam particles, a step that we describe in more detail below.

Now the number of secondary interactions is sampled from the probability distributions of Eq. (2.8) or Eq. (2.10) respectively. The chosen number of additional scatters is sampled according to the standard QCD $2 \rightarrow 2$ matrix elements and the same parton densities that were used for the hard process. That is, the additional hard processes are generated exactly according to the inclusive perturbative cross section, with no modification for the fact that they are additional scatterings. This list of processes is then successively processed by the parton shower. The partons involved in the additional hard scatters are also parton showered. As far as final state showering is concerned, this is identical to a standard hard process. For the initial state shower, we use the standard evolution algorithm, but with modified parton distribution functions, motivated by our model for extracting partons out of the hadron, which we return to shortly.

If the backward evolution of a scattering leads to a violation of four-momentum conservation, the scattering cannot be established. It is therefore regenerated until the desired multiplicity has been reached. If a requested scattering can never be generated without leading to violation of momentum conservation, the program eventually gives up, reducing the multiplicity of scatters.

After the parton shower, the quarks and gluons must be formed into the observed hadrons. The colour preconfinement property of the angular-ordered parton shower is used as the basis of the cluster model [44], which is used in **Herwig++** to model the hadronization. The cluster model however necessarily expects (anti)quarks or (anti)diquarks at the beginning of the hadronization. In the final state this prerequisite is easily fulfilled by the gluon splitting mechanism: all final-state gluons decay non-perturbatively to light quark–antiquark pairs. In the case of an initial-state parton from an incoming hadron, this necessitates a parton extraction model, which we describe in the next section.

Finally all unstable particles must be decayed. **Herwig++** uses a sophisticated model of hadronic decays as described in Refs. [47, 48].

2.2.1 Parton extraction

In the standard **Herwig++** treatment of a single hard scattering, the prerequisite that the outgoing partons must be (anti)quarks or (anti)diquarks is implemented by forcing the backward evolution to terminate on a valence parton. This then gives

a diquark as the proton remnant for example. This diquark is colour-connected through the colour connections of the valence quark either to a final-state parton emitted during the corresponding initial-state parton shower or through the hard process to a parton in one of the other jets in the event. In collisions other than pp , in events with little radiation, it can even be connected right through the event to the other hadron remnant.

It is often the case that by the time the perturbative evolution has terminated, the backward evolution has reached a valence parton, since their PDFs dominate at high x and low scale. When this is not the case and the backward evolution has terminated on a gluon or sea quark, one or two additional backward steps respectively are ‘forced’, using the standard backward evolution algorithm, but with all flavours except the one necessary for the forced step, vetoed.

In the implementation of multiple interactions, we keep the treatment of the first interaction untouched, i.e. it is exactly as just described. This means that the valence structure of the hadron has already been saturated, with one valence parton extracted and the remainder forming the hadron remnant. This does not therefore provide a structure that can be iterated for subsequent scatters. Instead, we modify the backward evolution so that it terminates on a gluon. We do this both dynamically during the evolution and by a forced backward evolution step if necessary. During the backward evolution we use modified parton distribution functions that are identical to the standard ones but with the valence contributions subtracted out¹. We stress that this subtraction of valence contributions is the only modification we make. In particular, the distribution of gluons is identical to that in the original hadron, leading to the possibility that the backward evolution of multiple scatters can over-saturate the available energy, which we deal with as already discussed above.

Once the backward evolution has terminated on a gluon, its colour connections can therefore be inserted into those of the previous remnant. As a concrete example, for the second scattering in an event with an incoming proton, the colour line of the gluon is connected to the diquark proton remnant and the anticolour line of the gluon is connected through the valence quark, to the outgoing parton that the diquark was previously connected to. This then gives a structure that can be iterated an arbitrary number of times. Since we do not order the additional hard scatters, for example in transverse momentum, this is equivalent to the colour connection model described as ‘random’ in [33]. The implementation of other colour connection models as described there would be possible, and may be interesting work for the future.

We illustrate this parton extraction model in more detail in Fig. 1. In the upper part of the figure, which is shaded, we can see the extracted partons after a possible perturbative parton shower. In the lower half of the figure, additional forced splittings

¹They do not therefore obey a momentum sum rule, but the algorithm is not sensitive to this fact, since it only involves ratios of PDFs. If one wanted to, one could rescale all the modified PDFs by a common factor to regain the momentum sum rule. The results would be unchanged.

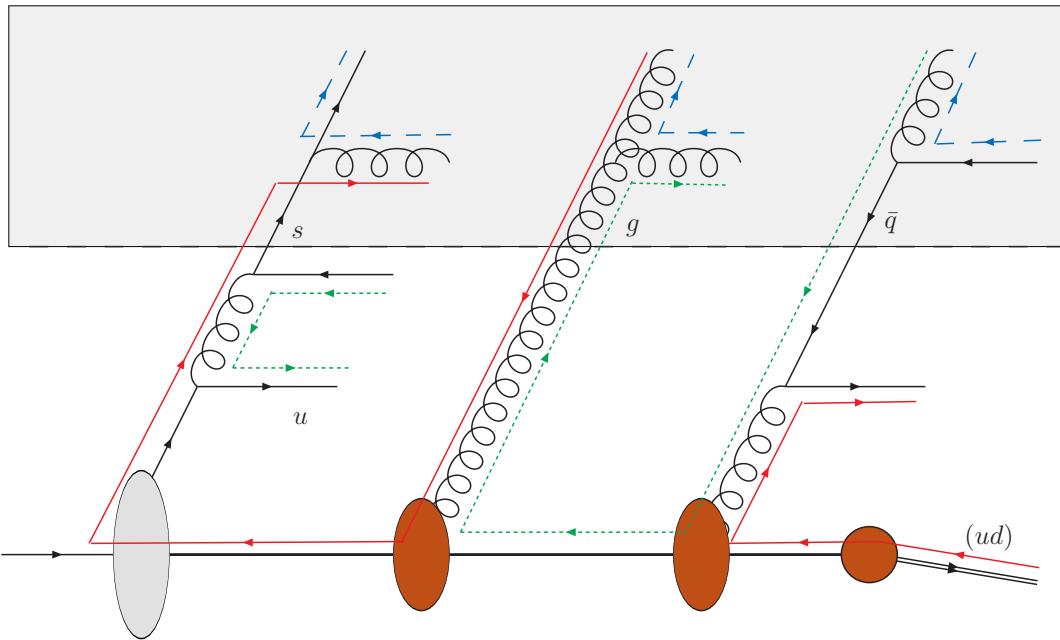


Figure 1: Schema of how the forced splittings and colour connections are implemented. Splittings in the shaded area stem from the hard scatters and the initial state parton shower. The final splittings at the bottom are non-perturbative.

are carried out in order to guarantee a certain flavour structure of the remnant. The first extracted parton will always be a valence quark while all additional hard scatters will always end up on a gluon. The colour structure is as just described, with the gluon produced by each hard scatter inserted into the colour-anticolour connection left by the previous one.

The way in which the structure of the hadron remnant is represented in the event record is not quite the same as the way in which it is generated, as described above. The same event is shown in Fig. 6 as it would appear in the event record, as described in Appendix B.

3. Results

We will now discuss several hadronic observables both for the Tevatron and the LHC. In particular a comparison to CDF data [26] is performed. For that reason the non-standard jet algorithm used for the data analysis has been implemented. Detector effects are solely taken into account by simulating the 92% track efficiency simply by ignoring 8% of charged particles, chosen randomly. For the LHC the prediction is compared to several other generators [35].

3.1 Tuning and Tevatron results

We have performed a tune of the model by calculating the total χ^2 against the

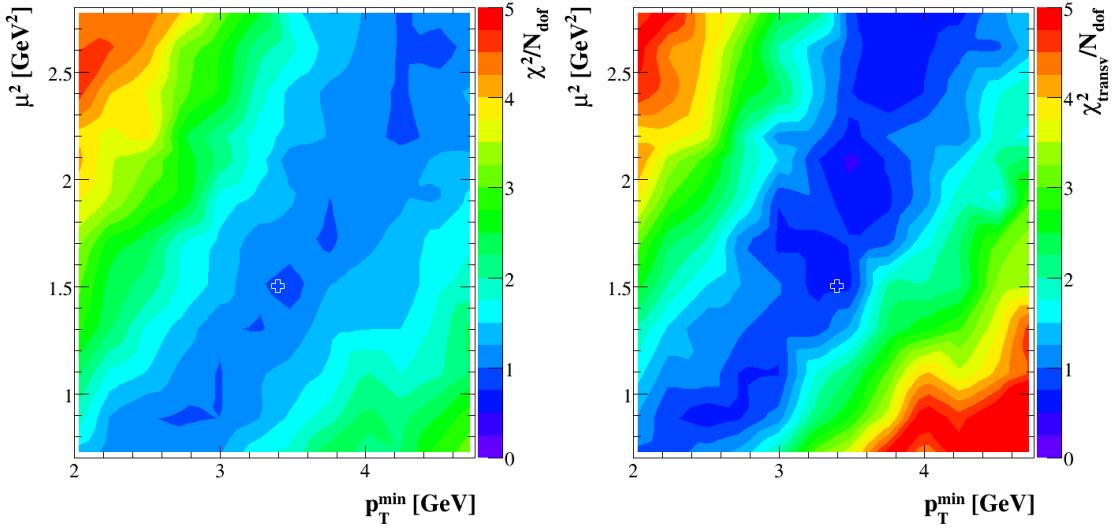


Figure 2: Contour plots for the χ^2 per degree of freedom of all discussed observables (left) and only the ones from the transverse region (right). The cross indicates the location of our preferred tune.

data from Ref. [26]. For this analysis each event is partitioned into three parts, the **towards**, **away** and **transverse** regions. These regions are equal in size in $\eta - \phi$ space and classify where particles are located in this space with respect to the hardest jet in the event. We compare our predictions to data for the average number of charged particles and for the scalar p_T sum in each of these regions. As we are aiming primarily at a good description of the underlying event in high p_T events, we used jet production with a minimal transverse momentum of 15 GeV as the signal process. Because of that we only use data for the region of leading jet transverse momentum above 20 GeV. We added an additional systematic error in quadrature for the lowest p_T bins as described in Appendix C.

The parameter space for this tune is two dimensional and consists of the p_T cutoff p_T^{\min} and the inverse hadron radius squared, μ^2 , entering $A(b)$ in Eq. (2.4). In Fig. 2 we show the contour plots for all six observables and for the observables from the transverse region respectively. We have used the MRST 2001 LO [49] PDFs built in to **Herwig++** for this plot, and discuss the PDF-dependence in the next section. For these, and all subsequent plots, we use **Herwig++** version 2.2.1, with all parameters at their default values except the two we are tuning and, in the next section, the PDF choice.

The description of the Tevatron data is truly satisfactory for the entire range of considered values of p_T^{\min} . For each point on the x -axis we can find a point on the y -axis to give a reasonable fit. Nevertheless an optimum can be found between 3 and 4 GeV. The strong and constant correlation between p_T^{\min} and μ^2 is due to the fact that a smaller hadron radius will always balance against a larger p_T cutoff as far as

the underlying event activity is concerned.

As a default tune we use $p_T^{\min} = 3.4$ GeV and $\mu^2 = 1.5$ GeV². Figure 7 shows the result of this parameter choice for the transverse region. The towards region is shown in Fig. 8 as well as the away region in Fig. 9. For these plots we used 10 million events in contrast to 1 million for each point in Fig. 2, which is the reason for the slight differences in the corresponding χ^2 values.

It is clear from these figures that event generation without any model for the underlying event is not capable of describing the data. In particular, in the transverse region, which receives the least contribution of the two jets from the matrix element, the results are a factor of two below the data.

Although our default multi-parton interaction (MPI) model gives a good overall description of the data, we see a slight trend to produce too much multiplicity in all the regions, most noticeably in the towards region, and too little p_T^{\sum} in all the regions, most noticeably in the away region. This corresponds to having a slightly too soft spectrum of individual particles and has also been observed in attempts to fit the fortran **HERWIG+Jimmy** model, the forerunner of ours, to the data of [26]. We note that in the towards region, which is dominated by the primary jet, **Herwig++** without MPI is already close to the data, leaving very little room for MPI effects. Almost any model of the underlying event will produce more than enough multiplicity here and overshoot the data. The same is true to a lesser extent in the away region. In the process of χ^2 minimization, there is therefore a slight pressure to suppress the underlying event effect, which results in the slight undershooting of the p_T^{\sum} predictions. The same effect is true even more weakly in the transverse region, where one would say that the description is very good, but there is a slight trend to be above the data for N_{ch} and below it for p_T^{\sum} . Since the effect is strongest for the regions dominated by the primary jets, we conclude that this is a general **Herwig** issue not specifically related to the MPI model. In any case, it is clear that it vastly improves the description of data relative to the no-MPI model.

We want to stress that the data from the experimental analysis are uncorrected. We already obtain a total χ^2 per degree of freedom very close to unity even with an over-simplified implementation of the reconstruction efficiency but a more precise examination would have to take detector effects into account in a more complete manner.

3.2 PDF uncertainties

For precision studies it is important to quantify the extent to which hard scattering cross sections are uncertain due to uncertainties in the PDFs. As we have already mentioned, jet cross sections are particularly sensitive to the amount of underlying event activity, which introduces an additional dependence on the PDF in our model. In particular, it relies on the partonic scattering cross sections down to small transverse momenta, which probe momentum fractions as small as $x \sim 10^{-7}$ at the LHC

and $x \sim 10^{-6}$ at the Tevatron, where the PDFs are only indirectly constrained by data. One will have measured the amount of underlying event activity at the LHC by the time precision measurements are being made, so one might think that the size of the underlying event correction will be known. However, in practice, jet cross section corrections depend significantly on rare fluctuations and correlations in the underlying event, so the correction must be represented by a model tuned to data, rather than by a single number measured from data. This will therefore entail in principle a retuning of the parameters of the underlying event model for each new PDF. This would make the quantification of PDF errors on a given jet cross section, or of extracting a new PDF set from jet data, much more complicated than a simple reweighting of the hard scattering cross section.

In this section we explore the extent to which this effect is important, by studying how the predictions with fixed parameters vary as one varies the PDF. We do this by comparing the central values of two different PDF sets (MRST and CTEQ) and also using the quantification of the uncertainties within one of them (CTEQ). Similar issues were also discussed in Ref. [50] for the uncertainty in parton shower corrections, which were found to be relatively small.

The results in Figs. 7–9 show the predictions of our model with MRST 2001 [49] and CTEQ6L [51] PDFs with the parameters fixed to the values obtained from our fit with the MRST PDFs. We see that the difference in the amount of underlying event activity, quantified by the results in the transverse region between 30 and 40 GeV as an example, is some 10% higher with CTEQ6L than with MRST.

To quantify the effect of the uncertainties within a given PDF set, we have used the error sets provided with the CTEQ6 family, and the formula

$$\Delta X = \frac{1}{2} \left(\sum_{i=1}^{N_p} [X(S_i^+) - X(S_i^-)]^2 \right)^{1/2}, \quad (3.1)$$

from Ref. [51]. Here X is the observable of interest and $X(S_i^\pm)$ are the predictions for X based on the PDF sets S_i^\pm from the eigenvector basis. Doing this naïvely, we found that the statistical error on independent runs with each PDF set was greater than the variation between the sets. To try to overcome this obstacle, we have studied the relative PDF uncertainty, i.e. $\Delta X/X(S_0)$, as a function of the number of points used for each $X(S_i^\pm)$.

As an example, we show the result in Fig. 3 for one bin corresponding to 35–36 GeV of the leading jet. The final statistics are obtained from 20M fully generated events for each PDF set and the value on the x axis is the number of events falling within this bin. We see that with these 20M events, we have still not completely eliminated the statistical uncertainties. However, a departure from the straight line on a log–log plot that would be expected for pure statistical errors, $\sim 1/\sqrt{N}$, is clearly observed. We use this to extract the *true* PDF uncertainty, by fitting a curve

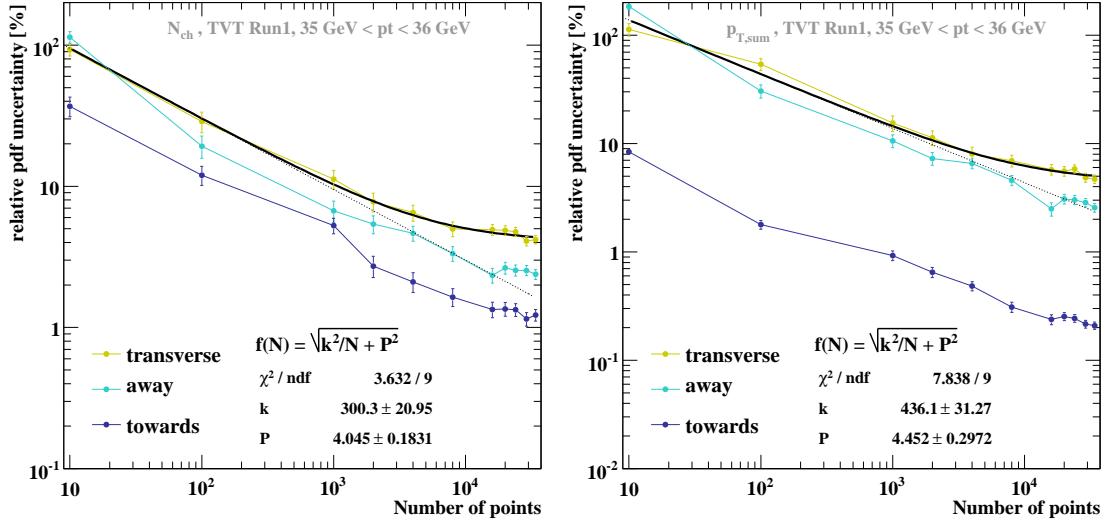


Figure 3: Relative PDF uncertainty, $\Delta X/X(S_0)$, in percent. Left for the multiplicity observables and right for the p_T^{sum} observables. The different curves show the results for the three different regions defined in the experimental analysis. The PDFs used are CTEQ6M [51] and its corresponding error sets. The fit result shown as a solid line is for the transverse region. Also shown as a light dashed line is the fit assuming a purely statistical error.

of the form

$$f(N) = \sqrt{\frac{k^2}{N} + P^2} \quad (3.2)$$

to these data. In performing the fit we get a reliable result already for a moderate number of events. From the fit results we can estimate the number of events that would be necessary to eliminate the contribution of the statistical uncertainty. Requiring it to be less than 10^{-1} of the total uncertainty leads to $N \sim 10^6$, which translates into $\sim 10^9$ fully generated events for each of the 40 PDF sets, which is not feasible in practice. Instead, using our fit, we have a clear indication that the PDF uncertainty is around 4% for the multiplicity and 4.5% for the p_T^{sum} in the transverse region.

It is note-worthy that the difference between the two PDF sets is larger than the uncertainty on each. Although, as we have already mentioned, the underlying event will have already been measured before making precision measurements or using jet cross sections to extract PDFs, a model tuned to that underlying event measurement will have to be used and its tuning will depend on the PDF set. We consider an uncertainty of 5–10% large enough to warrant further study in this direction.

3.3 LHC expectation

We start the discussion of our predictions for the LHC with the plots in Fig. 4, which are related to the total multiplicity and mean multiplicity flow in jet events. We show **Herwig++** with and without MPI. We used QCD jet production with a minimal p_T

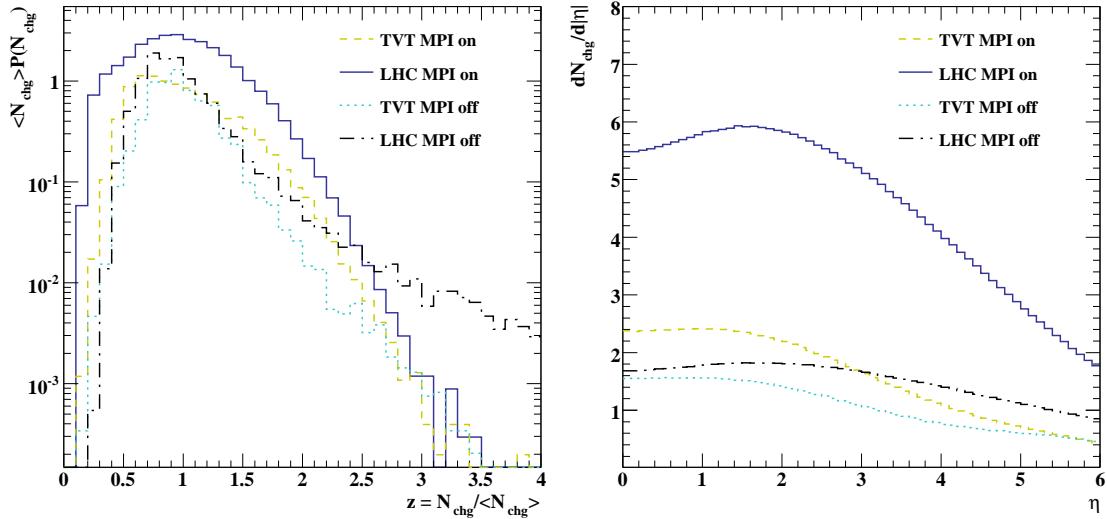


Figure 4: KNO plot (left) and differential multiplicity distribution (right) for Tevatron and LHC runs.

of 20 GeV as signal process. The MPI parameters were left at their default values, i.e. the fit to Tevatron CDF data.

The first plot in Fig. 4 shows the KNO distribution [52]. The MPI model satisfies KNO scaling fairly well, whereas `Herwig++` without an underlying event clearly violates it.

The second plot in Fig. 4 shows the mean charged multiplicity as a function of pseudorapidity, η . The effect of MPI is clearly visible, growing significantly from the Tevatron to the LHC.

In Ref. [35] a comparison of different predictions for an analysis modelled on the CDF one discussed earlier was presented. As a benchmark observable the charged particle multiplicity for the transverse region was used. All expectations reached a plateau in this observable for $p_T^{jet} > 10$ GeV. Our prediction for this observable is shown in Fig. 5, where it can be seen to have also reached a constant plateau within the region shown. The height of this plateau can be used for comparison. In Ref. [35] PYTHIA 6.214 ATLAS tune reached a height of ~ 6.5 , PYTHIA 6.214 CDF Tune A of ~ 5 and PHOJET 1.12 of ~ 3 . Our model reaches a height of ~ 5 and seems to be close to the PYTHIA 6.214 CDF tune, although our model parameters were kept constant at their values extracted from the fit to Tevatron data.

We have seen already in Sec. 3.1 that our fit results in a flat valley of parameter points, which all give a very good description of the data. We will briefly estimate the spread of our LHC expectations, using only parameter sets from this valley. The range of predictions that we deduce will be the range that can be expected assuming no energy dependence on our main parameters. Therefore early measurements could shed light into the potential energy dependence of the input parameters by simply

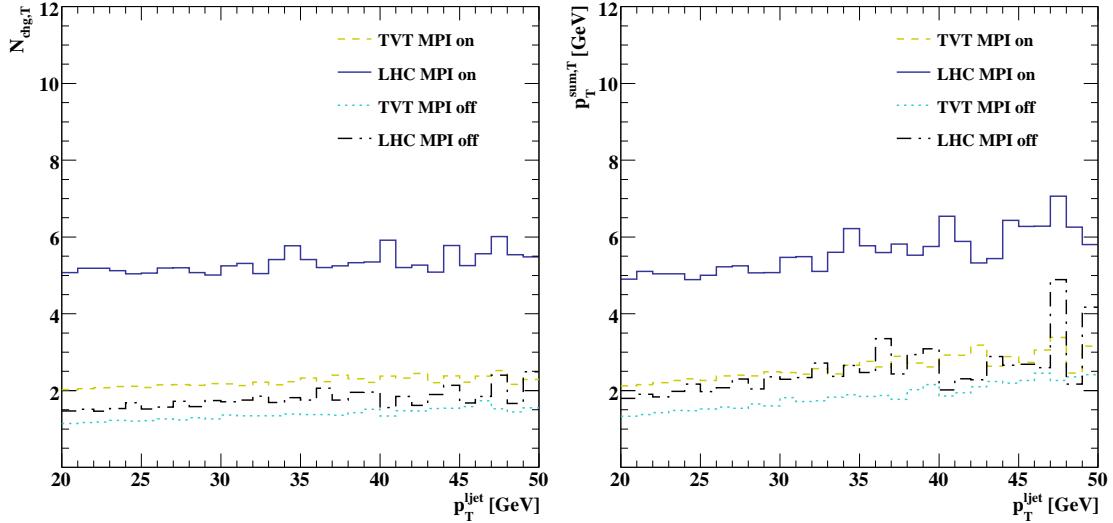


Figure 5: Multiplicity and p_T^{sum} in the transverse region for LHC runs with **Herwig++**. The different data sets are (from bottom to top): Tevatron with MPI off, LHC with MPI off, Tevatron with MPI on and LHC with MPI on.

comparing first data to these predictions. We extracted the average value of the two transverse observables shown in Fig. 5 for a given parameter set in the region $20 \text{ GeV} < p_T^{\text{jet}} < 30 \text{ GeV}$. We did that for the best fit points at three different values for p_T^{min} , namely 2 GeV, 3.4 GeV and 4.5 GeV.

LHC predictions	$\langle N_{\text{chg}} \rangle^{\text{transv}}$	$\langle p_T^{\text{sum}} \rangle^{\text{transv}} [\text{GeV}]$
TVT best fit	5.1 ± 0.3	5.0 ± 0.5

4. Conclusions

We have implemented a model of multiple parton interactions into the **Herwig++** event generator. We have tuned its free parameters to Tevatron data and found a good overall description. We have shown the extrapolation of its predictions to the LHC.

We consider the present work as only a first step towards our eventual goal of providing a complete description of the final state of minimum bias collisions and underlying events in hard hadron–hadron collisions, validated on and tuned to all available data and extrapolated to the LHC with quantified uncertainties.

Among the various phenomenological and theoretical studies that will be needed to achieve this goal, we mention the following avenues for future work. The present model only considers the contribution to multiple scattering from perturbative processes above p_T^{min} . We plan to extend it along the lines discussed in Ref. [20] to include non-perturbative partonic scattering below p_T^{min} . This will allow a description of minimum bias events, as well as the underlying event. There is a lot more

data available that constrains underlying event and minimum bias models to varying degrees. We plan to make a global analysis of this, in particular to give a handle on the energy dependence. It would also be interesting to consider whether the data require or allow an energy- (and scale-) -dependent effective proton radius, as predicted in Ref. [53]. Finally, it would be interesting to explore whether saturation effects are important and whether multiparton correlations, as discussed in [54], can be incorporated.

Acknowledgements

We would like to thank our collaborators on the **Herwig++** project for many useful discussions. Inspiring discussions with Jon Butterworth and the indispensable help of Leif Lönnblad are gratefully acknowledged. MB thanks the hospitality of the HEP group of UCL, where this study was completed. This work was supported in part by the European Union Marie Curie Research Training Network MCnet under contract MRTN-CT-2006-035606 and the Helmholtz–Alliance “Physics at the Terascale”. MB was supported by the Landesgraduiertenförderung Baden-Württemberg.

A. Forced splitting: implementation in the event record

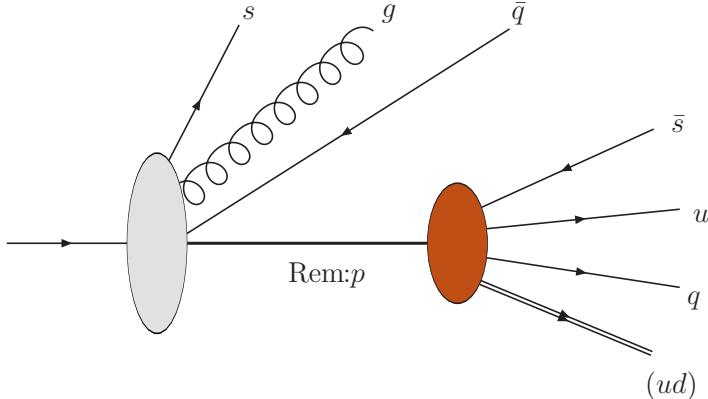


Figure 6: The structure of the event as it is implemented in **Herwig++**.

In Sect. 2 we have briefly described how the different hard scatters are correlated in colour space. This is of course an important model detail. In the event record, however, this will not be very obvious as this appears to be organized differently, in a way more closely related to the eikonal idea. In Fig. 6 we show the same particles (s, g, \bar{q}) that have already been extracted from the proton in the example of Fig. 1. This time the particles that have been extracted as last particles of the parton shower are directly extracted from the proton. All additional emissions of partons that are

related to the forced splitting, described in Sect. 2 appear as decay products of the intermediate remnant. In this way we emphasize the non-perturbative origin of these partons and draw a clear line between the perturbative parton shower model and the non-perturbative mechanism of forced splittings in the event record.

B. Model parameters

In the `Herwig++` manual [5], the general mechanism for accessing and changing parameters and switches of models is described, together with the main parameters and switches of the underlying event model. For completeness, we repeat the latter here.

Cuts: Via a cuts object the minimal p_T of the additional scatters can be set. This is one of the two main parameters of the model. The current default, obtained from the fit to Tevatron data described above, is 3.4 GeV.

InvRadius: The inverse beam particle radius squared. The current default is 1.5 GeV², obtained from the above mentioned fit.

Algorithm: A switch to enable efficient generation of additional scatters in rare (high- p_T) signal processes. Steers whether to use Eq. (2.8) or Eq. (2.10). The options are:

- 0: Underlying event process and signal process are identical.
- 1: Underlying event process and signal process are of the same type but the signal cross section is small. Here a veto algorithm has to be applied, if an additional scatter is produced with p_T larger than the cutoff for the hard process.
- 2: Underlying event process and signal process are distinct scattering types and the signal cross section is small. This is the default choice.

C. Systematic errors in the low p_T region

When making the initial comparison with data, we observed a $> 3\sigma$ discrepancy for the observable $p_{T,sum}^{low}$ below 30 GeV of the leading jet. Above 30 GeV, this discrepancy is completely absent. However, we have almost no freedom to tune this observable, because it is completely dominated by the p_T of the jet itself. For the same reason, the relative error is extremely small in this region, $\sim 0.5\%$, so the absolute discrepancy is only about 2%. Nevertheless if we are going to fit to data in this region, we need to understand this effect, to avoid the χ^2 of the fit being completely dominated by it.

From Ref. [26] we find that the data sample was obtained by requiring a calorimeter tower with $E_T > 20$ GeV (including charged and neutral particles), described as the ‘Jet20’ sample. The analysis however is based on charged particle tracks. In particular the x -axis in all observables is the scalar p_T sum of the charged particles defined to be in the hardest jet. It is clear that this sample is only unbiased for large enough values of p_T^{jet} relative to the 20 GeV trigger. Where this happens however is not obvious. In Ref. [26] the sample was assumed to be perfectly unbiased from 20 GeV onwards. This statement is based on the good match between the Jet20 data and the Min Bias sample around that value. Any judgement on the smoothness of the match is however limited by the statistical error on the Min Bias data, which is becoming large in the region where the two samples overlap. Therefore we have added an additional systematic error in quadrature to the data points to reflect the precision with which we are confident they are unbiased. We choose this to have the form

$$\sigma_{add} = \frac{\sigma_{sys}^0}{10} \left(30.5 - \frac{p_T}{\text{GeV}} \right) \text{ for } (20.5 < p_T/\text{GeV} < 30.5) , \quad (\text{C.1})$$

where σ_{sys}^0 is extracted from the uncertainties in the bins 18 – 21 GeV of the Min Bias data and the linear form ensures that the additional uncertainty goes to zero for $p_t \sim 30$ GeV. In more detail, we extract σ_{sys}^0 by fitting these three bins with a linear function and use the uncertainty on the value at 20.5 GeV from this fit for σ_{sys}^0 (in practice, this procedure gives only a slightly smaller error than simply averaging the errors on the three bins). This gives the following values for the $p_{T,sum}$ observables:

region	σ_{sys}^0
towards	440 MeV
away	1950 MeV
transverse	840 MeV

For the multiplicities we obtain the following values:

region	σ_{sys}^0
towards	0.75
away	1.07
transverse	0.63

References

- [1] G. Corcella *et. al.*, *HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010, [[hep-ph/0011363](#)].
- [2] S. Gieseke, A. Ribon, M. H. Seymour, P. Stephens, and B. Webber, *Herwig++ 1.0: An event generator for e^+e^- annihilation*, *JHEP* **02** (2004) 005, [[hep-ph/0311208](#)].

- [3] S. Gieseke *et. al.*, *Herwig++ 2.0 release note*, [hep-ph/0609306](#).
- [4] M. Bähr *et. al.*, *Herwig++ 2.1 release note*, [arXiv:0711.3137](#).
- [5] M. Bähr *et. al.*, *Herwig++ Physics and Manual*, [arXiv:0803.0883](#).
- [6] M. Bähr *et. al.*, *Herwig++ 2.2 release note*, [arXiv:0804.3053](#).
- [7] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026, [[hep-ph/0603175](#)].
- [8] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to PYTHIA 8.1*, [0710.3820](#).
- [9] T. Gleisberg *et. al.*, *SHERPA 1.α, a proof-of-concept version*, *JHEP* **02** (2004) 056, [[hep-ph/0311263](#)].
- [10] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063, [[hep-ph/0109231](#)].
- [11] J. Alwall *et. al.*, *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J.* **C53** (2008) 473–500, [[0706.2569](#)].
- [12] **UA5** Collaboration, G. J. Alner *et. al.*, *The UA5 High-Energy $\bar{p}p$ Simulation Program*, *Nucl. Phys.* **B291** (1987) 445.
- [13] T. Sjöstrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, *Phys. Rev.* **D36** (1987) 2019.
- [14] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Phys.* **C72** (1996) 637–646, [[hep-ph/9601371](#)].
- [15] J. M. Butterworth and M. H. Seymour, “Jimmy4 Manual.” Downloadable under <http://projects.hepforge.org/jimmy/>.
- [16] A. Capella, U. Sukhatme, C.-I. Tan, and J. Tran Thanh Van, *Jets in Small $p(T)$ Hadronic Collisions, Universality of Quark Fragmentation, and Rising Rapidity Plateaus*, *Phys. Lett.* **B81** (1979) 68.
- [17] A. Capella and J. Tran Thanh Van, *A New Parton Model Description of Soft Hadron-Nucleus Collisions*, *Phys. Lett.* **B93** (1980) 146.
- [18] A. Capella and J. Tran Thanh Van, *Hadron - Nucleus Interactions and the Leading Particle Effect in a Dual Parton Model*, *Z. Phys.* **C10** (1981) 249–262.
- [19] R. Engel, *Photoproduction within the two component dual parton model. 1. Amplitudes and cross-sections*, *Z. Phys.* **C66** (1995) 203–214.
- [20] I. Borozan and M. H. Seymour, *An eikonal model for multiparticle production in hadron hadron interactions*, *JHEP* **09** (2002) 015, [[hep-ph/0207283](#)].

- [21] **Axial Field Spectrometer** Collaboration, T. Akesson *et. al.*, *Double parton scattering in pp collisions at $\sqrt{S} = 63 \text{ GeV}$* , *Z. Phys.* **C34** (1987) 163.
- [22] J. Pumplin, *Hard underlying event correction to inclusive jet cross sections*, *Phys. Rev.* **D57** (1998) 5787–5792, [[hep-ph/9708464](#)].
- [23] **CDF** Collaboration, F. Abe *et. al.*, *Study of four jet events and evidence for double parton interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8 \text{ TeV}$* , *Phys. Rev.* **D47** (1993) 4857–4871.
- [24] **D0** Collaboration, V. M. Abazov *et. al.*, *Multiple jet production at low transverse energies in $p\bar{p}$ collisions at $\sqrt{s} = 1.8 \text{ TeV}$* , *Phys. Rev.* **D67** (2003) 052001, [[hep-ex/0207046](#)].
- [25] **CDF** Collaboration, F. Abe *et. al.*, *Double parton scattering in $\bar{p}p$ collisions at $\sqrt{s} = 1.8 \text{ TeV}$* , *Phys. Rev.* **D56** (1997) 3811–3832.
- [26] **CDF** Collaboration, A. A. Affolder *et. al.*, *Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV*, *Phys. Rev.* **D65** (2002) 092002.
- [27] **CDF** Collaboration, D. E. Acosta *et. al.*, *The underlying event in hard interactions at the Tevatron $\bar{p}p$ collider*, *Phys. Rev.* **D70** (2004) 072002, [[hep-ex/0404004](#)].
- [28] A. Moraes, C. Buttar, and I. Dawson, *Prediction for minimum bias and the underlying event at LHC energies*, *Eur. Phys. J.* **C50** (2007) 435–466.
- [29] L. Fano, *Minimum bias and underlying event at CMS*, *Acta Phys. Polon.* **B38** (2007) 435–442.
- [30] D. Acosta *et. al.*, *The underlying event at the LHC*, . CERN-CMS-NOTE-2006-067.
- [31] I. Borjanovic *et. al.*, *Investigation of top mass measurements with the ATLAS detector at LHC*, *Eur. Phys. J.* **C39S2** (2005) 63–90, [[hep-ex/0403021](#)].
- [32] C. Hackstein, “Simulation of final states in vector boson fusion.” Diploma thesis, Institut für Theoretische Physik, Universität Karlsruhe, Nov. 2007.
- [33] T. Sjöstrand and P. Z. Skands, *Multiple interactions and the structure of beam remnants*, *JHEP* **03** (2004) 053, [[hep-ph/0402078](#)].
- [34] T. Sjöstrand and P. Z. Skands, *Transverse-momentum-ordered showers and interleaved multiple interactions*, *Eur. Phys. J.* **C39** (2005) 129–154, [[hep-ph/0408302](#)].
- [35] S. Alekhin *et. al.*, *HERA and the LHC - A workshop on the implications of HERA for LHC physics: Proceedings Part A*, [hep-ph/0601012](#).
- [36] S. Catani, M. Ciafaloni, and F. Hautmann, *High-energy factorization and small x heavy flavor production*, *Nucl. Phys.* **B366** (1991) 135–188.

- [37] J. C. Collins and R. K. Ellis, *Heavy quark production in very high-energy hadron collisions*, *Nucl. Phys.* **B360** (1991) 3–30.
- [38] E. M. Levin, M. G. Ryskin, Y. M. Shabelski, and A. G. Shuvaev, *Heavy quark production in parton model and in QCD*, *Sov. J. Nucl. Phys.* **54** (1991) 867–871.
- [39] S. Höche, F. Krauss, and T. Teubner, *Multijet events in the $k(T)$ -factorisation scheme*, 0705.4577.
- [40] J. Bartels, M. Salvadore, and G. P. Vacca, *AGK cutting rules and multiple scattering in hadronic collisions*, *Eur. Phys. J.* **C42** (2005) 53–71, [[hep-ph/0503049](#)].
- [41] L. Durand and P. Hong, *QCD and Rising Cross Sections*, *Phys. Rev. Lett.* **58** (1987) 303.
- [42] L. Durand and H. Pi, *Semihard QCD and high-energy pp and $\bar{p}p$ scattering*, *Phys. Rev.* **D40** (1989) 1436.
- [43] S. Gieseke, P. Stephens, and B. Webber, *New Formalism for QCD Parton Showers*, *JHEP* **12** (2003) 045, [[hep-ph/0310083](#)].
- [44] B. R. Webber, *A QCD Model for Jet Fragmentation including Soft Gluon Interference*, *Nucl. Phys.* **B238** (1984) 492.
- [45] G. Marchesini and B. R. Webber, *Simulation of QCD Jets including Soft Gluon Interference*, *Nucl. Phys.* **B238** (1984) 1.
- [46] G. Marchesini and B. R. Webber, *Monte Carlo Simulation of General Hard Processes with Coherent QCD radiation*, *Nucl. Phys.* **B310** (1988) 461.
- [47] D. Grellscheid and P. Richardson, *Simulation of Tau Decays in the Herwig++ Event Generator*, 0710.1951.
- [48] D. Grellscheid, K. Hamilton, and P. Richardson, “Simulation of Meson Decays in the Herwig++ Event Generator.” in preparation.
- [49] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *MRST2001: Partons and alpha(s) from precise deep inelastic scattering and Tevatron jet data*, *Eur. Phys. J.* **C23** (2002) 73–87, [[hep-ph/0110215](#)].
- [50] S. Gieseke, *Uncertainties of Sudakov form factors*, *JHEP* **01** (2005) 058, [[hep-ph/0412342](#)].
- [51] J. Pumplin *et. al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [[hep-ph/0201195](#)].
- [52] Z. Koba, H. B. Nielsen, and P. Olesen, *Scaling of multiplicity distributions in high-energy hadron collisions*, *Nucl. Phys.* **B40** (1972) 317–334.

- [53] R. M. Godbole, A. Grau, G. Pancheri, and Y. N. Srivastava, *Soft gluon radiation and energy dependence of total hadronic cross-sections*, *Phys. Rev.* **D72** (2005) 076001, [[hep-ph/0408355](#)].
- [54] T. C. Rogers, A. M. Stasto, and M. I. Strikman, *Unitarity Constraints on Semi-hard Jet Production in Impact Parameter Space*, [0801.0303](#).

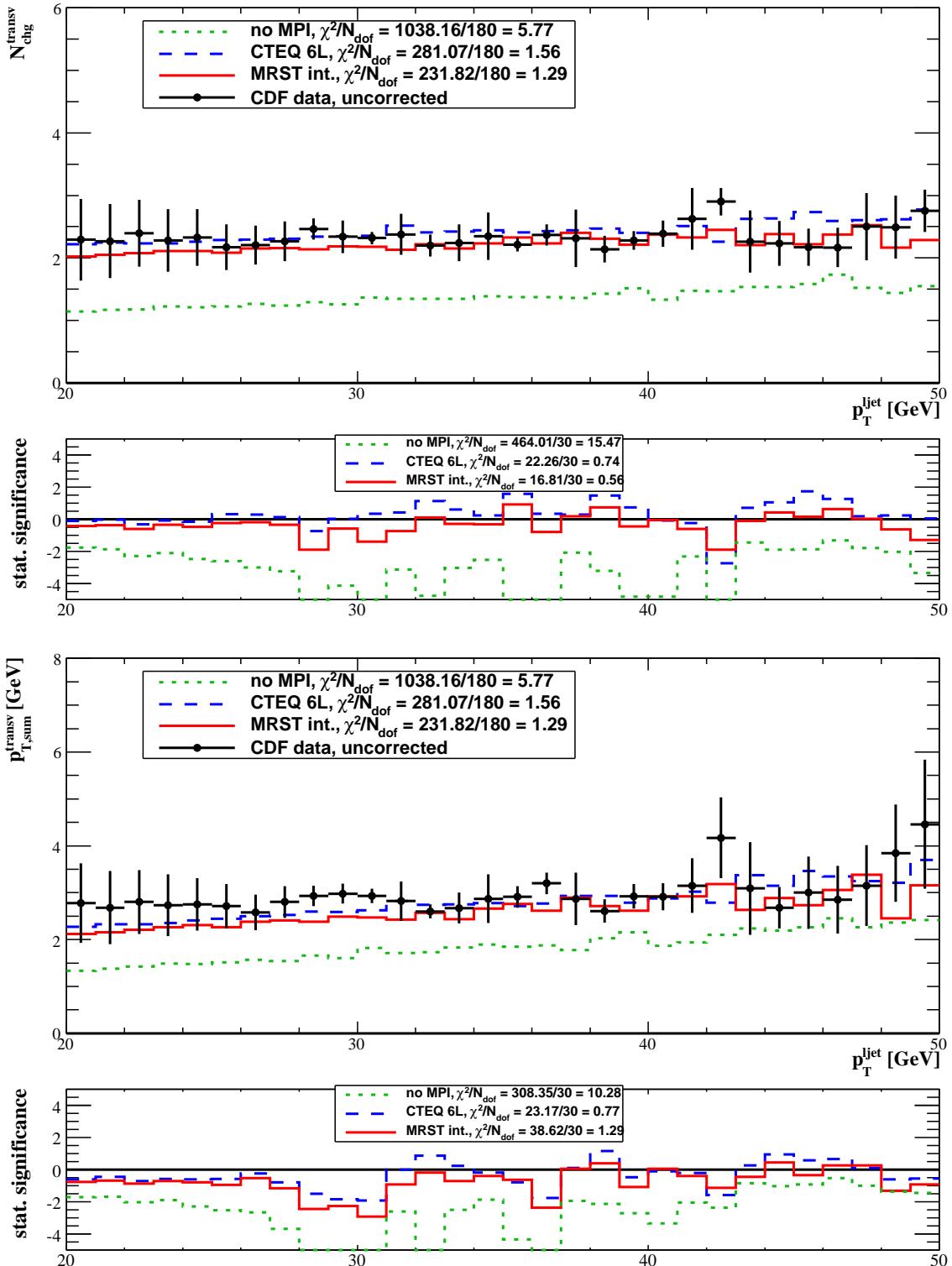


Figure 7: Multiplicity and ptsum in the **transverse** region. CDF data are shown as black circles. Herwig++ without MPI is drawn in green dots, Herwig++ with MPI using MRST [49] PDFs in solid red and with CTEQ6L [51] as blue dashed. The lower plot shows the statistical significance of the disagreement between Monte Carlo prediction and the data. The legend on the upper plot shows the total χ^2 for all observables, whereas the lower plot has the χ^2 values for this particular observable.

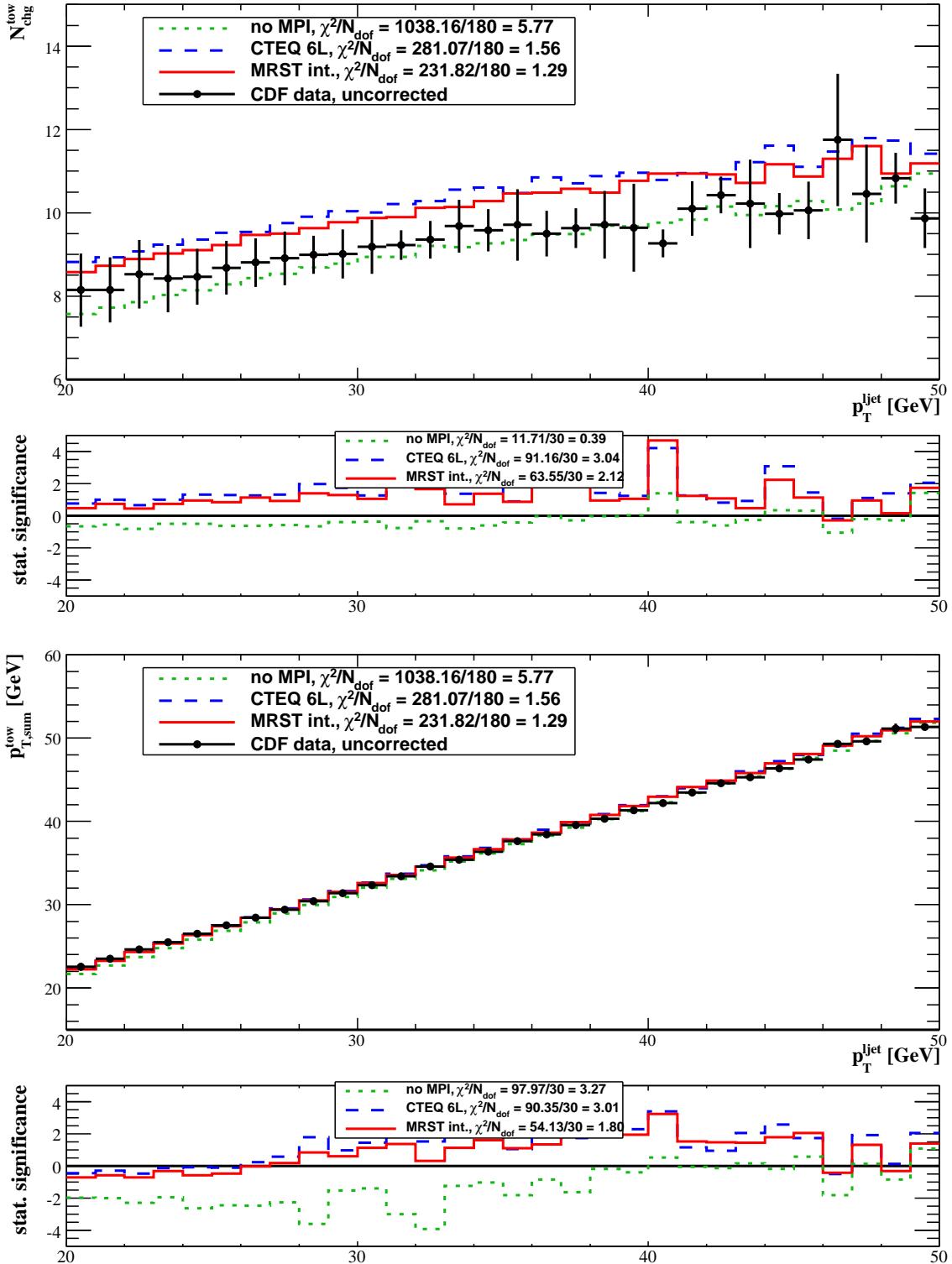


Figure 8: Multiplicity and p_{tsum} in the **towards** region. CDF data are shown as black circles. **Herwig++** without MPI is drawn in green dots, **Herwig++** with MPI using MRST [49] PDFs in solid red and with CTEQ6L [51] as blue dashed. The lower plot shows the statistical significance of the disagreement between Monte Carlo prediction and the data. The legend on the upper plot shows the total χ^2 for all observables, whereas the lower plot has the χ^2 values for this particular observable.

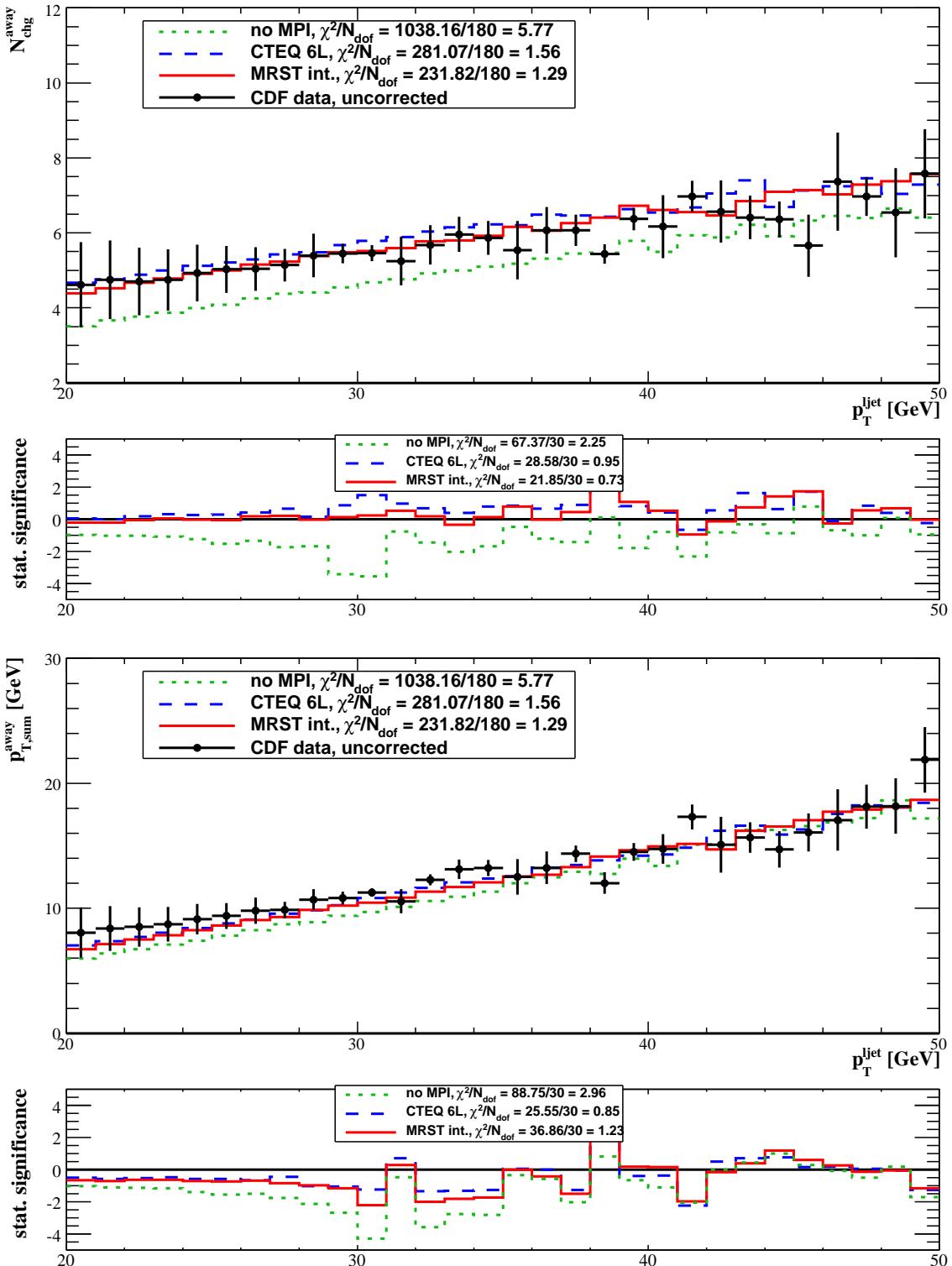


Figure 9: Multiplicity and ptsum in the **away** region. CDF data are shown as black circles. Herwig++ without MPI is drawn in green dots, Herwig++ with MPI using MRST [49] PDFs in solid red and with CTEQ6L [51] as blue dashed. The lower plot shows the statistical significance of the disagreement between Monte Carlo prediction and the data. The legend on the upper plot shows the total χ^2 for all observables, whereas the lower plot has the χ^2 values for this particular observable.